AFRL-AFOSR-VA-TR-2017-0014

SATA STOCHASTIC ALGEBRAIC TOPOLOGY AND APPLICATIONS

Shmuel Weinberger
UNIVERSITY OF CHICAGO THE
5801 S ELLIS AVE
CHICAGO, IL 606375418

01/23/2017
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/RTA2

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 12-29-2016	**2. REPORT TYPE** Final Report	**3. DATES COVERED** *(From - To)* 09/30/2011-9/29/2016

4. TITLE AND SUBTITLE
SATA STOCHASTIC ALGEBRAIC TOPOLOGY AND APPLICATIONS

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-11-1-0216

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Shmuel Weinberger (PI)
Yuliy Baryshnikov (CoPI)
Jonathan Taylor (CoPI)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Chicago, 5801 S Ellis Ave, Chicago, IL 60637
Stanford University, 340 Panama St, Stanford, CA 94305
University of Illinois Urbana-Champaign, 506 S. Wright Street, Urbana IL 61801

8. PERFORMING ORGANIZATION REPORT NUMBER
FP047558

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
875 North Randolph Street
Suite 325, Room 3112
Arlington VA, 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
UChicago

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION A: Distribution approved for public release.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This project was devoted mainly to applications of topology primarily to data analysis, but also to some engineering (e.g. control) problems. Because of noise and uncertain environments, stochasticity is an important element. Topological invariants are robust to some errors in the bulk, but can frequently be highly sensitive to outliers. The work done in this project concerns the amount of data necessary to solve topological inference, even free of noise, and also the nature of errors caused by noise: Different kinds of tail behavior have very different implications, and heavy tails are shown to have severe implications for these methods. Also studied is how much data is necessary to compute topological invariants robustly as a complexity theoretical problem and also as an analysis of algorithms problem, and under what kinds of conditions of local featurelessness of the data (sometimes called a condition number or feature size)? The study of critical points is applied to using these methods for inference within machine learning, and the topology of configuration spaces is applied to control problems. Finally, several of the papers studied nontraditional integrals (Euler integration) which are related to the Gaussian Kinematic Formula, and have earlier been used for target enumeration,

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU		Shmuel Weinberger
					19b. TELEPHONE NUMBER *(Include area code)* 773-702-7349

Reset

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

DISTRIBUTION A: Distribution approved for public release.

AFOSR Award # FA9550-11-1-0216

SATA STOCHASTIC ALGEBRAIC TOPOLOGY AND APPLICATIONS

Final Report

December 2016

PI.
Shmuel Weinberger
Department of Mathematics
University of Chicago

CoPIs

Yuliy Baryshnikov
Departments of Mathematics and Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

And

Jonathan Taylor
Department of Statistics
Stanford University

Table of Contents

Summary

This report summarizes the five year SATA (Stochastic Algebraic Topology and Applications) project involving three US investigators together with Robert Adler and his team at the Technion (Israel), who is working on a similar grant with the European Office of Aerospace Research and Development. We have not always been able to separate the work done by the American group from the Israeli team.

The project has led to approximately 50 papers, the vast majority of which are already published or submitted. These include results related to the statistics of random functions, random complexes, and random manifolds and embeddings. In addition, there have been some extensions of the scope of the project to include some new problems and areas related to the theme (stochastic algebraic topology) that were not mentioned in the original proposal in terms of connections to e.g. order statistics, sample complexity bounds, and topological sampling theory.

This grant has also played a role in the training of graduate students and postdoctoral fellows, dissemination of results and general educational activity on the importance of stochastic algebraic topology as tool in data analysis.

Objectives.

Recall that the Stochastic Algebraic Topology and Applications (SATA) project aims to exploit recent advances in the complementary areas of topology and stochastic processes to tackle a wide range of data analytic problems of broad importance. Treating data topologically is crucial in scenarios in which it is important to detect, localize, and perhaps perform an initial classification of objects without attempting to completely characterize them. Adding a stochastic element allows for the almost pervasive situation in which the data itself is imperfectly observed due to the presence of background noise. As current probabilistic and statistical methodology is ill suited to detect such qualitative structures, the project aims to develop generic stochastic models whose topological structures are amenable to mathematical analysis, as a first step towards implementation of a broader, more quantitative program. Core topics include random functions on manifolds, random manifolds created by random embeddings, and random manifolds arising in machine learning, along with their theoretical and practical interplay. Secondary topics include the analysis of associated algorithms, and the topological understanding of random spaces that arise in particular stochastic models. We have also studied implementation and application of these ideas on some problems coming from engineering and physics.

Accomplishments

This is basically a mathematics-based project, and the methods are hard analysis supplemented with, and often motivated by, computation. Below we describe three general areas in which we have made significant progress, and a fourth area of new ideas that were generated by our projects. The headings all correspond to topics in the original proposal, which provides background material.

The research in this project grew out of two different types of topological data analysis: statistical and geometric. The statistical problems were from the theory of Gaussian (related) random fields, and use topological invariants as proxies for measurements of more direct interest (such as excursion probabilities) and also can be used as robust signatures of complicated data (such as time series, random fields, moving objects, etc.)[1]. The second source was geometric -- trying to find useful geometries of data that can be used to improve data analysis[2].

The topological sampling problem is one of the first theoretical problems in topological data analysis: assuming that data is being sampled from a probability

[1] A convenient source for this is the book “Random Fields and Geometry” by Adler and Taylor and the draft of a second volume (coauthored with K. Worsley) that can be found on Adler’s web page.

[2] An overview of the philosophy of this work can be found in G. Carlsson’s survey paper in Acta Numerica (Vol 23 (2014) 289-368.)

distribution of geometric origin, can one determine the underlying geometry? The literature on this and related problem has grown substantially and the picture is considerably different than it was at the beginning of the project.

At the time the project began, a number of papers showed that if a distribution was sampled from a "well conditioned" manifold with "not too much" noise, then with "enough" samples, with high probability, one could recover the topology (homotopy type or even the diffeomorphism type) of the manifold. Different and deep versions of this result, and new algorithms that work more or less efficiently in different contexts, are still being discovered and progress on this basic problem remains important.

Among the issues that our work has illuminated directly related to topological sampling are:

(1) The theoretical limits on when topological reconstruction is possible [20].
(2) The rates of convergence of specific algorithms for computation of topological invariants (and related phase transitions) such as homology and homotopy [4,5,17,18] or Euler integrals [30].
(3) Connections to percolation theory [51].
(4) The behavior of these algorithms on pure noise (as preparation for understanding better noisy data sets) [4,50,31].
(5) Stability properties of topological invariants of functions [16].
(6) Lower bounds on sampling, Kolmogorov and logical complexity of some of these reconstruction problems [47].
(7) What happens to (6) in the generic, as opposed to worst case, situations [7]. ([7] directly solves one of the main problems identified in the original proposal, and we will discuss it in more detail below.)
(8) Reparametrization invariant functionals on time series [52,53].
(9) Typical shapes of discretized loops with topological constraints [8].
(10) Applications of critical point theory to statistical inference problems [12,19,21,24,25,26,27,28,34,35,36,37,38,40,41].

We also made progress in furthering the applications of this work, developing new topological signatures [44, 53], and have initiated work on finding some new topological invariants that are more computable than the traditional ones. Finally, we mention some results that stem from our topological study that are not particularly stochastic.

Below we describe three general areas in which we have made significant progress and a fourth area of new ideas that are spurred by our projects. The headings all correspond to topics in the original proposal for ease of comparison, which provides background material.

1: Statistics of random functions

From a mathematical standpoint, our basic setup here starts with a base topological space, M, typically but not always a manifold, and a random function f on M, with values in R^n for some $n \geq 1$. Given a realization of such a function, we proposed studying various statistics pertaining to it, including properties of the set of critical points, critical values, sub- and super-level sets, etc. These arise in both themes mentioned above.

(i) In joint work of Adler and Taylor with Eliran Subag, a Technion graduate student, they provided a new approach, along with extensions, to results in two important papers of Worsley, Siegmund and coworkers closely tied to the statistical analysis of fMRI (functional magnetic resonance imaging) brain data. These papers studied approximations for the exceedence probabilities of scale and rotation space random fields, the latter playing an important role in the statistical analysis of fMRI data. The techniques used there came either from the Euler characteristic heuristic or via tube formulae, and to a large extent were carefully attuned to the specific examples of the paper. In [1] they treated the same problem, but via calculations based on the so-called Gaussian kinematic formula. This allowed for extensions of the Worsley-Siegmund results to a wide class of non-Gaussian cases. In addition, it allows one to obtain results for rotation space random fields in any dimension via reasonably straightforward Riemannian geometric calculations. Previously only the two-dimensional case could be covered, and then only via computer algebra.

(ii) The paper by Adler, Moldovskaya and Samorodnitsky [2] studied, in a one dimensional setting, the problem of whether or not two or more points which lie in an excursion set of a smooth random process belong to the same connected component. This is a fundamental problem, at the level of connectivity, that has eluded successful analysis for a number of years.

Adler and Samorodnistky, in a later paper [6], take this much further, to the setting of continuous Gaussian random fields on higher dimensional Euclidean spaces, and address the question of how likely it is for the excursion sets to have a ``hole" of a certain dimension and depth? Answering this question in full generality appears to be impossible at the moment, but their paper makes significant progress. Specifically, they determine how likely is such a field to be above a high level on one compact set (e.g. a sphere) and to be below a fraction of that level on some other compact set, (e.g. at the center of the corresponding ball). These questions have clear, and sometimes surprising and counter-inuitive, answers at the level of large deviations.

(iii) Naitzat and Adler [30] proved a central limit theorem for the Euler integral of a Gaussian random field. Recall that Euler integrals of deterministic functions have recently been shown to have a wide variety of possible applications, including in signal processing, data aggregation and network sensing. Adding random noise to these scenarios, as is natural in the majority of applications, leads to a need for statistical analysis, the first step of which requires asymptotic distribution results for estimators. The first such result is provided in this paper, as a central limit theorem for the Euler integral of pure, Gaussian, noise fields.

Proving these results turned out to be somewhat more complicated than was originally expected. Fortunately, the actual central limit theorem is simple to state and, equally importantly, simple to apply as an inference tool in real life scenarios. On the other hand, the proof required a sophisticated manipulation of the most recent advances in central limit theorems for Gaussian functionals based on representations via the Malliavin calculus.

This work assumes additional significance because of the work of Baryshnikov and Ghrist relating Euler integrals to approaches to target enumeration. These authors and Wright [14] develop a general Hadwiger theory for these invariants, which helps explain their centrality.

(iv) Baryshnikov and Weinberger have studied the generic behavior of persistent homology in various function spaces [16]. Unlike the usual stability theorems that assert that long bars are stable with respect to small perturbations, the nature of the results of this paper describe the stability of the small bars with respect to smooth but possibly large perturbations (where here large means large magnitude). We call this kind of information, the “jitter” of a function or a space. For a simple example, f(x) + sin nx will have many local minima and maxima for n large, whenever f is a Lipschitz function on an interval, say [a,b]. (And this number grows like n(b-a), i.e. linearly in n and the length.)

These statistics can give information about underlying mechanisms for data and the paper discusses sizes of craters on the moon and stock price time series (where the persistent homology is consistent with a Holder ½ function -- although more subtle dendrogram type invariants do distinguish these from a time reparametrised exponentiated Brownian motion [52, 53]). These and additional theoretical results about singularities and mathematical analogues are being written into a revised version of the paper on jitter. It also has a large-scale aspect, as well, and is related to the short paper [47].

Note the strange nature of the integrand (note that intervals appear in it![3]): it is essentially a current describing the average number of times one should see an interval approximately of length [0, x/2] in the persistence diagram. The blow up near the origin is because, with many points, there are very many very short persistence intervals. Its quadratic nature is perhaps typical. *Similar effects occur for Brownian motion, as mentioned above*. The nonzero measure associated to long intervals is essentially a precise form of the crackle phenomenon.

(v) Baryshnikov [52,53] initiated a study of reparametrization-invariant functionals of time series, an important addition to the standard toolbox of data anlysis, almost entirely relying on harmonic analysis, e.g. Fourier transform in its different avatars.

One direction deals with the realization that the Reeb tree that can be associated to a scalar univariate function carries more information than merely its barcode: the bars have a chirality, i.e. can fall or raise. Statistics of these raising or falling bars can, signify in a reparametrization invariant way the asymmetry of the process. In [52] Baryshnikov studies the baseline case of various Brownian motions, and the resulting asymmetries.

[3] Some might prefer the intervals replaced by characteristic functions of the intervals, and then this formula can be viewed as an equality in a space of measures.

On the empirical side, he exhibited time irreversibility in several classes of time series Another model considered by Baryshnikov deals with cyclic but not necessarily periodic processes (such as business cycles). Motivated by a classical theorem of K.-T. Chen, he introduced an algorithm to recover a *cyclicity* [53], a family of time series sequentially following a similar periodic pattern, perhaps with a desynchronized clock. The algorithm relies on the notion of iterated integrals, and leads to remarkably reliable reconstructions of cyclic orderings in some systems, in particular in the relative performance of industrial sectors during the business cycles of the US economy.

2: Morse theory, critical points, Betti numbers and random complexes

As described in the original proposal, we planned to invest considerable effort in the study of the topological properties of random simplicial complexes for hypothesis testing.

Adler and Bobrowski [5] considered for a finite set of points P in R^d the behavior of the number of critical points of the distance function $d_P : R^d \rightarrow R^+$ which measures Euclidean distance to the set P. In particular, they studied the number of critical points of d_P when P is a random sample from a given distribution, and the limit behavior of N_k = the number of critical points of d_P with Morse index k, as the number of points in P goes to infinity. They gave explicit computations for the normalized, limiting, expectations and variances of the N_k, as well as distributional limit theorems. These results are related to recent results of Kahle in which the Betti numbers of the random Cech complex based on P were studied. The practical implication of these results lies in the design of sampling algorithms for manifold learning via approximating simplicial complexes.

Similar ideas, applied to a different regime, were used by Bobrowski and Weinberger [18] to discover the phase transitions in the computation of homology of Riemannian manifolds from Cech complexes, at least in the case of flat tori. They gave a heuristic indicating that the same results should apply in general, but did not give quantitative results about the rate of convergence. This is ongoing work.

On another front, Adler and Yogeshwaran [50] have studied random complexes (generally Cech or Rips) generated from point clouds in settings where the underlying point process is neither Poisson nor a simple random sample, but comes from a general stationary process in which there may be considerable correlations (either positive or negative) between different regions. A typical example of significant current interest from both theoretical and applied points of view is given by determinantal point processes. Their surprising finding is that many of the results from the better known, and simpler, scenarios, while they do carry over in principle to the correlated situation, involve quantitative differences which are going to be important in any learning or estimation scenario.

Additional results have been written up in [51] for which the limit regime is in the so-called `thermodynamic' regime (which includes the percolation threshold) in which the complexes become very large and complicated, with complex homology characterised by diverging Betti numbers. The proofs combine probabilistic arguments from the theory of stabilizing functionals of point processes and topological arguments exploiting the properties of Mayer-Vietoris sequences. The Mayer- Vietoris arguments are crucial, since homology in general, and Betti numbers in particular, are global rather than local phenomena, and most standard probabilistic arguments are based on the additivity of functionals arising as a consequence of locality. These results are closely related to the ideas about the role of topological testability and are one of the main future directions of this work.

A related problem on which Adler, Bobrowski and Weinberger have made considerable progress [4] is a phenomenon that we have named `crackle'. Once again, random Cech complexes are created, this time with fixed inter-point distances and based over different types of samples. It was shown that if the additional noise is in some sense large then sample points can appear basically anywhere, introducing extraneous homology elements. We observe that Gaussian noise does not crackle (which explains why topological methods have been of most use in that sample) but exponential and scale-free has a lot of crackle.

A family of results, including a law of large numbers, will appear in a joint work of all the PI’s and co-PI’s. Among these is the result that, for a sample of n exponential variables, the expected number of bars in the zero-th order persistence homology of length in the interval [x,y] tends, as n tends to infinity, to

$$\int e^{-u}(1-e^{-u})^{-2}\ 1_{[2x,2y]}(u)\ du.$$

The blow up near the origin is results from the fact that, with many points, there are very many very short persistence intervals. The quadratic nature of the divergence is perhaps typical. Similar things occur for persistence intervals in the level set filtrations of Brownian motion. Overall, the nonzero measure associated to long intervals is essentially a precise formulation of the crackle phenomenon.

In a subsequent paper [17] far more precise results are established. There, point process convergence of spherically symmetric k-tuples $(Xi_1 ,...,Xi_k)$ ☒ (R^d) is studied under certain geometric constraints. If the law of the random points in R^d has either regularly varying or exponentially decaying tails that vanish slowly enough, then a certain Poisson random measure becomes the weak limit of the point process. On the contrary, if the law of the random points has rapidly decaying exponential tails, the corresponding point process tends to zero in probability. As an application, the homology of the Cech complex built over those random points is studied. The weak convergence result shows that Betti numbers of order up to d -1 have either Poisson limits or are degenerate, depending upon how heavy are the tails of the distributions of the random points.

As this paper was being written up it became clear that although the work was originally motivated by, and has immediate applications to, topological data analysis, the main results in some sense “belong” to the classical area of extreme value theory (EVT). Consequently, the paper was written in the language of EVT, for two reasons. The first was one of convenience - most of the natural notation and a lot of the needed pre-existing lemmas came from there. The second was more long term, in that the authors wanted to introduce to the EVT community, in a language with which they are familiar, the general area of stochastic algebraic topology.

Despite this, the paper concludes in the language of applied topology, and allows the topology literate reader to understand the implications of EVA type analysis to applied topology.

Owada has taken this further in subsequent papers [32,33]. While still motivated by the issue of crackle in TDA, and thus interested in the behavior of Betti numbers and other topological aspects of Cech complexes, the approach taken in this work is to investigate the limiting behavior of a sub-graph counting process, when the graph in question is the 1-skeleton of the complex. In particular, the subgraph counting process at the core of the paper counts the number of subgraphs having a specific shape that exist outside an expanding ball as the sample size increases. As an underlying law, the paper considers distributions with a regularly varying tail and those with an exponentially decaying tail.

The aim is then to obtain *functional* limit theorems for these processes as the underlying scale parameter of the Cech complex changes. This is, of course, a much more sophisticated result than a standard (central) limit theorem, and is, to the best of our knowledge, the first time that a functional limit theorem has been proven in the setting of random topology. Regarding the specific results, it is seen that the nature of the functional central limit theorem differs according to the speed at which the ball expands. In fact, the proper normalizations for the limit theorems and the properties of limiting Gaussian processes are all determined by whether or not an expanding ball covers a region - called a weak core - in which the random points are densely scattered and form a giant geometric graph.

The results of this work not only have implications for increased understanding of the structure of persistent homology under crackle - an issue of applied relevance - but have significant intrinsic interest. In particular, the limiting stochastic processes that appear here seem to be completely new in the context of probability theory.

3: Random manifolds and random embeddings

In the initial proposal we noted that random manifolds arise in a number of scenarios, and that one of the key geometric quantities that arises there in recovering the homology of a manifold $M \subset R^n$ by randomly sampling points from it is the critical radius τ of the manifold.

Roughly speaking, the reach, or critical radius, of a manifold is a measure of its departure from convexity that incorporates both local curvature and global topology. It plays a major role in many aspects of differential geometry, and more recently has turned out to be a crucial parameter in assessing the efficiency of algorithms for manifold learning.

As the critical radius depends on the embedding it is of interest to study the behavior of the critical radius of a Riemannian manifold (M; g) for a generic, or random embedding of M into R^n for large n. A natural model to consider is based on taking independent, identically distributed, copies, f_1,..., f_n of a real-valued random field on M and then working with $f = (f_1,...,f_n) : M \rightarrow R^n$ to define the random, embedded manifold f(M). Each such random embedding gives rise to a random Riemannian metric on M, that is naturally related to the original metric g. Other geometric features of interest include the study of the geometric invariants of such Riemannian metrics, such a volume, curvature, diameter, etc.

Together with a Technion graduate student, Sreekar Ram Krishnan, Adler, Taylor and Weinberger have shown [7] that the self-normalised critical radii of these randomly embedded manifolds converges almost surely to a deterministic limit determined by the structure of the underlying manifold M and the covariance function of the process
Somewhat unexpectedly, this limit turns out to be the same one that arises in studying the exceedance probabilities of the Gaussian process over the manifold.

En passant it was also proven that the induced embeddings are asymptotically isometric, from which it follows that other properties of the embedded manifolds, such as volume, curvature integrals, etc, also converge. Underlying this there turns out to be a much deeper notion of convergence. [23] proves such results.

This collection of theorems results in an important philosophical implication, which is encouraging for topological data analysis. Although the sample complexity of learning a manifold grows (exponentially) with ambient dimension (see [47]), even with lower bounds on the critical radius and upper bound on diameter, for a given “platonic ideal”, the random embeddings do not suffer this defect, and generically the image manifold can be learned with a sample complexity that does not grow with dimension even in the presence of (controlled) noise.

4. Other directions that have grown out of this work.

(i) The lower complexity bounds in the problems, established, in general, in [47] pose an important issue for TDA. Many of the usual questions people ask are unfeasible in general: computation of invariants is too difficult, the number of topological types is too large. Applications of topological methods must either explain why the data should be suitable for those methods - e.g. why the complications that could arise, do not (as in the work [7] mentioned in the previous paragraph NOT MENTIONED THERE) - or they should be focused on invariants that can be measured. Weinberger has been studying such invariants, modeled on testability properties of graph properties. The simplest of these is the Euler characteristic divided by the volume - which is essentially (for Riemannian manifolds) the average (Pfaffian of the) curvature. As an average, it is subject to sampling. Thus, a large submanifold in Euclidean space whose average curvature is large will surely have complicated topology, and discovering its topological properties will require enormous sampling and computational resources.

Similarly, characteristic 0 Betti numbers seem to be testable (but not too straightforwardly: random regular graphs have high Euler characteristic and first Betti number: but randomly they look like trees that have no local topology! (The fact that the ratio can be approximated via finite samples is a consequence of Hodge theory). Whether this is the case for mod p Betti numbers is an important problem.

Weinberger has shown that for "geometric complexes" i.e. those which embed in Euclidean space with bounded geometry, the phenomenon of "foamless foam", i.e. of the presence of large homology without any homology being present at a small scale, does not occur. This suggests that better algorithms for computing such invariants might be possible for the geometric case. These results are consistent with the results of [51] on thermodynamic limits. Thermodynamic limits also arise in the dual question: throwing away the many visible small cycles and look for the birth of the large, macroscopic cycles. Bobrowski and Weinberger have been studying this and its connection to percolation theory. These works are not yet complete.

(ii) Another "spin off" is the paper [20] which deals with the question of whether there are manifolds that can be made arbitrarily close to one another in Gromov-Hausdorff space with a local contractibility function. Any such manifolds must be homotopy equivalent and must have the same rational characteristic classes. (The first is easy; the second is a deep theorem of S. Ferry.)

However, we show that there are indistinguishable manifolds, and even some infinite sets of such manifolds. We also show that for "reasonable fundamental groups" the set of doppelgangers of a given manifold is finite. However, there are some.

In revised versions of this paper, the connection between this topological problem and analytic methods based on C* algebras has been strengthened, resulting in the paper being rewritten to take this into account. As a consequence the theory is now essentially complete for many fundamental groups (including abelian, and torsion free linear groups).

(iii) Taylor and several co-authors have been studying inferential problems in statistics and machine learning related to critical points of common objective functions encountered in machine learning. A canonical example of such an objective function would the LASSO (squared error plus an L^1 penalty). The solution to this problem is a critical point, and many of the tools developed in the theory of smooth random fields on piecewise smooth spaces developed by Adler and Taylor are applicable to such problems.

Taylor and several co-authors have continued work on selective inference reported in SATA's 2014 annual report. The main methodological contribution [21] describes a formal approach to inference after model selection where model selection is broadly described as observing partial information about the entire sample. Previously reported work [25] applied an early version of this framework to inference after selecting features using the LASSO.

One of the key constructions in [13] is the idea of performing inference under a selected model as opposed to inference for different parameters of the same model as in [14]. This technical distinction allows for valid inference in regression models with unknown variance [27]. Other work in this vein include a version of selective inference in which the sample is randomized before a model selection algorithm is applied [40]. These algorithms are similar to those appearing in the differential privacy literature (c.f. “The reusable holdout: preserving validity in adaptive data analysis", Science, 2015). The intervals and hypothesis tests in [40] are less conservative than the differential privacy approach.

Work is continuing on applying this approach to model selection from a continuum of models as in [19] in which a selective inference algorithm is proposed for a sequential algorithm to determine the rank of spiked covariance model in PCA. The proposed algorithm is less conservative than asymptotic approaches based on RMT. Current work is focused on applying these techniques to CCA (canonical correlations analysis) and relating our finite sample size algorithms to the asymptotic RMT approaches. The complete development of these ideas appear in the papers [12,19,21,24,25,26,27,28, 34, 35,36,37, 38, 40, 41]

[37] builds on the exact selective inference of [25] in the Gaussian least squares setting to the setting of general likelihoods with a LASSO penalty. It describes how to remove the parametric modelling assumption for the covariance, using a bootstrap estimate of covariance, removing the assumption that the selected model is correct.

The model in [26] considers selective inference in regression problems where features are clustered and one uses a prototype to represent the entire cluster in a regression model. They derive an analog of the F-test for the entire cluster given its prototype was selected in a model selection procedure like the LASSO.

[41] Building on the framework of selective inference after randomization in [37], we describe a simple randomization scheme that yields an explicit formula for the selective likelihood ratio which is necessary for selective inference. The construction relies on an exact inversion of the KKT conditions of a particular randomized LASSO problem. Co-author Nan Bi was supported by AFOSR in carrying out this work.

[54] This paper extends the approach of [41] to fairly arbitrary convex programs. Notably, penalties with some curvature such as the group LASSO can easily be handled in this fashion, as well as multiple steps forward stepwise and \top K" marginal screening. For penalties with curvature, the change-of-measure formula involves a Jacobian encoding similar geometric structure to the Jacobian in Steiner-Weyl volume-of-tubes formulae. [34] This paper considers selective inference in a Bayesian context, building on an approach for univariate problems in [49]. The main technical difficulty in this work is computing the selection probability as a function of the parameter on which a prior is specified. We use a large-deviations approximation to this probability that involves solving a well-defined convex program for each step of the Markov Chain. This program possesses nice structure, particularly if selection is randomized. [21] produces a sequential model selection algorithm that satisfies the hypotheses of the sequential FDR-controlling procedure of [22]. We demonstrate an improvement in power over the spacings test of [41] whose tests also fail to satisfy the hypotheses required for control of FDR.

(iv) Baryshnikov and Mileyko have been studying problems related to many of the above, but on networks. The persistence homology of networks (with analytic growth bounds, with respect to steadily increasing scale) has been related to network flow and congestion problems as well as to defining dimensions of networks[4]. Baryshnikov and his postdoc Yuri Mileyko continued the studies of the (local) dimensions of synthetic and real-life networks. The background for this quest was an emerging trend in networking community to view networks as hyperbolic in some sense (e.g. as being CAT(0) spaces, or Gromov-hyperbolic, etc). One thread in this area of research had as an underlying premise that the real-life networks can be properly modeled by random geometric graphs sampled from a ball in homogeneous hyperbolic spaces, or, even more specifically, from the hyperbolic plane. While the pictorial representations in the numerous publications in this spirit looked convincing, the basic questions were not asked, i.e. why the hyperbolic plane? Why the assumptions of homogeneity? etc.

In general, the random finite metric space obtained by a dense enough sampling from a Riemannian manifold would provide enough data to detect at least the dimension of the underlying manifold: if $X \subset M$ is a finite sample from M, a manifold of dimension m, then for spherical shells of points in X, and judiciously chosen radii R, r (R much less than the injectivity radius, r large to ensure dense sampling), the persisting homology of the Rips complex of SX(x; R; r) = H(X, X-x; R,r) should be concentrated in dimensions 0 and m, for interior points of the sample, and just in dimension 0 for the points near the boundary.

Experiments confirmed that this is exactly what happens for the samples from hyperbolic plane. However, contrary to what one might expect from the existing literature, the analysis of the ASN network (the world-visible structure of the autonomous domains, roughly the network of Internet connections) shows that their local homologies are extremely wild and irregular, and are nowhere close to the sample from the hyperbolic plane (or any manifold). On the positive side, the local homology is yet another characteristic of the nodes in large graphs, and we plan to use it systematically for network analysis (and, perhaps, to analyze samples from singular spaces, in a TDA fashion).

The results of these experiments are available at the web site http://publish.illinois.edu/ymb/2014/09/21/dimension-of-the-internet/ and show how the local homologies behave for samples of the hyperbolic plane and for the “internet graph”.

Other applications of these methods large networks obtained from sampling large geometric function spaces will appear in the revised version of [16]

(v) S. Mukherjee and Katharine Turner have developed a persistent homology transform that has application to shape statistics – describing a shape by a “Radon transform”” of the persistent homology of the height functions in all the different directions, and applied this, together with D. Boyer in the Evolutionary Anthropology Department at Duke University, to data comparing calcanei bones of various primates

[4] See the paper of Block and Weinberger on Large Scale Homology theories of Metric spaces and Baryshnikov, Bonahon, Jonckheere and Lou, on Euclidean versus Hyperbolic congestion for some background, all available on the authors’ web pages.

[33]. Turner and V. Robins have been studying persistence homology invariants of sand and other disordered materials [34]. This, too, has strong connections with theory of testable invariants mentioned above.

(vi) Arnold, Baryshnikov and Mileyko's paper [25] studies the typical shapes of the (discretized) loops sampled uniformly from the space of loops in a given (free) homotopy class on a surface. It is shown that there exists a large deviation principle forcing the sampled loop to be close to the solution of a variational problem. (For example, in the case of plane with punctures, to a collection of straight segments representing the minimal loop in the given homotopy class.) We note that this result is in tension with the large amount of time it can take loops that are near to an index zero geodesic that is not actually length minimizing. A given homotopy class can have infinitely many such geodesics even for a bumpy metric on S^2. Nevertheless, asserts [25] in that case, "almost all" geodesics will be "pointlike".

(vii) In [9], Baryshnikov studies problems related to tiling spaces -- a topic closely related to the mathematical physics of testability and to the problem of defining invariants that can be computed quickly. The classical Wang (2D)-tilability problem asks whether one can tile the plane using a collection of domino tiles (with marked boundaries, under the matching boundaries constraints). Motivated by some questions from Markov Random Fields, we investigate same problem under constraints on the (asymptotic) frequencies of tiles of each type. There are some natural conditions coming from matching the boundary frequencies, but, as it turned out, they are not sufficient. We prove that the realizable frequencies form a convex proper subset of the polyhedron of feasible frequencies. In a sequel (finalized now with Abram Magner and Spankowski) we ask for more general question: what is the average asymptotic genus of a 2D surface with a free Z2 action admitting a tiling with given frequencies of tiles.

(viii) We conclude with discussing some more engineering applications of topological ideas. These were done by Baryshnikov and collaborators.

In the three papers [11, 13, 46] the authors analyze the topology of the configuration space from the viewpoint of complexity of any feedback control stabilizing the trajectory under the (stochastic or not) perturbations on an attractor. The topology of the configuration space is critical for this structure of the feedback control loop. In the first paper, we look at the configuration space of multi-legged robotic device that turned out to be related to moment-angle complexes and Stanley-Reisner rings.

It is well-known that randomly switching between multiplications by several operators can lead to divergent dynamics, even if each operator in the family is asymptotically contracting. Some algebraic conditions (such as solvability of the Lie algebra generated by the operators) prevent such anomalies. Solvability is not an open property, motivating the study [15] , that proves that slightly relaxing the solvability condition keeps the switched systems stable.

Personnel supported at some point during the grant period

Shmuel Weinberger, PI,
Department of Mathematics,
University of Chicago

Yuliy Baryshnikov, co-PI
Departments of Mathematics and Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

Jonathan Taylor, co-PI
Department of Statistics
Stanford University

Partially supported the related joint work of and with:
Robert Adler
Department of Electrical Engineering
Technion

Nan Bi (graduate student)
Statistics
Stanford

Cheng Chen (graduate student)
UIUC

Yunjin Choi, (graduate student)
Statistics
Stanford

Sunder Ram Krishnan (graduate student)
Electrical engineering
Technion

Joshua Loftus (graduate student)
Statistics
Stanford

Abraham Magner (postdoc)
Engineering
UIUC

Yuriy Mileyko (postdoc)
Mathematics
UIUC

Gregory Naitzat (graduate student)

Electrical engineering
Technion

Yonatan Rosmarin (graduate student)
Electrical Engineering
Technion

Emily Schlafly (graduate student)
UIUC

Eliran Subag (graduate student)
Electrical Engineering
Technion

Takashi Owada (graduate student)
Electrical engineering
Technion

Xiaoying Tian (graduate student)
Statistics
Stanford

Katharine Turner (Graduate Student)
Mathematics
University of Chicago

Han Wang (Graduate Student)
Mathematics
University of Illinois at Urbana-Champaign

Elad Yarkoni (graduate student)
UIUC

Yogeshwaran Dhandapani (Postdoc)
Electrical Engineering
Technion

Publications.

1. R.J. Adler, E. Subag and J.E. Taylor, Rotation and scale space random fields and the Gaussian kinematic formula. *Annals of Statistics*, 40, 2012, 2910–2942.

2. R.J. Adler, G. Samorodnitsky and J.E. Taylor, High level excursion set geometry for non-Gaussian infinitely divisible random fields. *Annals of Probability*, 41, 2013, 134–169.

3. R.J. Adler, E. Moldavskaya and G. Samorodnitsky, On the existence of paths between two points in high level excursion sets of Gaussian random fields. *Annals of Probability*, 42, 2014, 1020–1053.

4. R.J. Adler, O. Bobrowski and S. Weinberger, Crackle: The homology of noise. *Discrete and Computational Geometry*, 52, 2014, 680–704.

5. O. Bobrowski and R.J. Adler, Distance functions, critical points, and topology for some random complexes. *Homology, Homotopy and Applications*, 16, 2014, 311–344.

6. R.J. Adler and G. Samorodnitsky, Climbing down Gaussian peaks. *Annals of Probability*, 2016. In press. (33 pages)

7. R.J. Adler, S.R. Krishnan, J.E. Taylor and S. Weinberger, Convergence of the reach for a sequence of Gaussian-embedded manifolds. Submitted for publication. (50 pages)

8. M. Arnold, Yu. Baryshnikov, Yu. Mileyko, Typical representatives of free homotopy classes in a multi-punctured plane, 2015, submitted.

9. Yu. Baryshnikov, J. Duda, W. Szpankowski, Types of Markov Fields and Tilings, IEEE Trans. on Information Theory, 2016.

10. Yu. Baryshnikov, V. Blumen, K. Kim, V. Zharnitsky Billiard dynamics of bouncing dumbbell, Physica D: Nonlinear Phenomena 269, 21-27, 2015.

11. Yu. Baryshnikov, B. Shapiro, How to Run a Centipede: a Topological Perspective In: Geometric Control Theory and sub-Riemannian Geometry, Springer INdAM Series 5, 37-51, 2014.

12. Xiaoying, Tian, Harris et al. Selective sampling after solving a convex problem". arXiv:1609.05609 [math, stat] (Sept. 2016). arXiv: 1609.05609.

13. Baryshnikov, Yuliy; Chen, Cheng; Wang, Han,A design of hybrid feedback stabilization on 1D coverage with topological perspectives,"American Control Conference (ACC), 2016,5154-5160,2016,American Automatic Control Council (AACC)

14. Y. Baryshnikov, R. Ghrist, M. Wright Hadwiger's Theorem for Definable Functions, Adv. Math. 245, 573-586, 2014.

15. Y. Baryshnikov, Liberzon, Daniel,Robust stability conditions for switched linear systems: Commutator bounds and the Łojasiewicz inequality,52nd IEEE Conference on Decision and Control,722-726,2013,IEEE

16. Y. Baryshnikov and S.Weinberger, Persistence Jitter and Texture (under revision)

17. O.Bobrowski, S.Mukherjee, and J.Taylor, Topological Consistency via Kernel Estimation, arXiv:1407.5272 [math, stat] (July 2014) arXiv: 1407.5272. to appear in Bernoulli

18. O.Bobrowski and S.Weinberger, On the vanishing of homology in Random Cech complexes. Random Structures and Algorithms, published online 7 November 2016

19. Yunjin Choi, Jonathan Taylor, and Robert Tibshirani. Selecting the number of principal components: estimation of the true rank of a noisy matrix. arXiv:1410.8260 [stat], October 2014. arXiv: 1410.8260.

20. A.Dranishnikov, S.Ferry, and S.Weinberger, An infinite dimensional phenomenon in finite dimensional topology (submitted)

21. William Fithian, Dennis Sun, and Jonathan Taylor. Optimal Inference After Model Selection. arXiv:1410.2597 [math, stat], October 2014. arXiv: 1410.2597.

22. Max Grazier G. Sell et al. Sequential selection procedures and false discovery rate control". en. In: Journal of the Royal Statistical Society: Series B (Statistical Methodology 78.2 (Mar. 2016), pp. 423-444.

23. S.R. Krishnan, J.E. Taylor and R.J. Adler, The intrinsic geometry of some random manifolds. Electronic Communications in Probability, in press. (13 pages)

24. Jason D Lee, Jonathan E Taylor, Exact Post Model Selection Inference for Marginal Screening (submitted)

25. Jason D. Lee, Dennis L. Sun, Yuekai Sun, Jonathan E.   , “Exact post-selection inference with the lasso”. Annals Statistics 44.3 (June 2016), pp. 907-927.

26. Stephen Reid, Jonathan Taylor, Robert Tibshirani, Post-selection point and interval estimation of signal sizes in Gaussian samples, In: Journal of the American Statistical Association (Oct. 2016),

27. Xiaoying Tian, Joshua R. Loftus, and Jonathan E. Taylor. Selective inference with unknown variance via the square-root LASSO. arXiv:1504.08031 [math, stat], April 2015. arXiv: 1504.08031.

28. Joshua R. Loftus and Jonathan E. Taylor. A significance test for forward stepwise model selection. arXiv:1405.3920 [stat], May 2014. arXiv: 1405.3920.

29. E Munch, K Turner, P Bendich, S Mukherjee, J Mattingly, J Harer Probabilistic Fréchet means for time varying persistence diagrams Electronic Journal of Statistics 9, 1173-1204

30. G. Naitzat and R.J. Adler, A central limit theorem for the Euler integral of a Gaussian random field. *Stochastic Processes and its Applications*, 2016. In press. (32 pages)

31. T. Owada and R.J. Adler, Limit theorem for point processes under geometric constraints (and topological crackle) *Annals of Probability*, 2016. In press. (54 pages)

32. T.Owada, Functional central limit theorem for subgraph counting processes (preprint)

33. T. Owada, Functional central limit theorem for subgraph counting processes, *Annals of Probability*, 2016. In press. (35 pages)

34. Snigdha Panigrahi, Jonathan Taylor, and Asaf Weinstein. Bayesian Post-Selection Inference in the Linear Model". In: arXiv:1605.08824 [stat] (May 2016). arXiv: 1605.08824.

35. Jonathan Taylor, Joshua Loftus, Ryan Tibshirani, Tests in adaptive regression via the Kac-Rice formula (submitted)

36. Jonathan Taylor, Richard Lockhart, Ryan J. Tibshirani, Robert Tibshirani, Postselection adaptive inference for Least Angle Regression and the Lasso (submitted)

37. Jonathan Taylor and Robert Tibshirani. Post-selection inference for L^1-penalized likelihood models. In: Canadian Journal of Statistics To appear (Feb. 2016).

38. Xiaoying Tian, Nan Bi, and Jonathan Taylor. MAGIC: a general, powerful and tractable method for selective inference". In: arXiv:1607.02630 [math, stat] (July 2016). arXiv: 1607.02630.

39. G. Thoppe, D. Yogeshwaran, and R.J. Adler, On the evolution of topology in dynamic clique complexes. *Advances in Applied Probability*, 48, 2016. In press. (33 pages)

40. Xiaoying Tian and Jonathan E. Taylor. Selective inference with a randomized response". In: arXiv:1507.06739 [math, stat] (July 2015). arXiv: 1507.06739.

41. Ryan J. Tibshirani et al. Exact Post-Selection Inference for Sequential Regression Procedures". In: Journal of the American Statistical Association 111.514 (Apr. 2016), pp. 600-620.

42. K.Turner, Cone fields and topological sampling in manifolds with bounded curvature, Journal of FoCM **13** (2013), no. 6, 913–933.

43. K Turner, Y Mileyko, S Mukherjee, J Harer , Fréchet means for distributions of persistence diagrams Discrete & Computational Geometry 52 (1), 44-70

44. Katharine Turner, Sayan Mukherjee, Doug M Boyer, Persistent homology

transform for modeling shapes and surfaces, Information and Inference, iau 11 2014

45. K.Turner and V.Robins, Principal Component Analysis of Persistent Homology Rank Functions with case studies of Spatial Point Patterns, Sphere Packing and Colloids (preprint)

46. Wang, Han, Chen, Cheng, Baryshnikov, Yuliy, A Topological Perspective on Cycling Robots for Full Tree Coverage, Algorithmic Foundations of Robotics XI, 659-675, 2015,Springer

47. S.Weinberger, The complexity of some basic topological inference problems, Journal of FoCM, 14 (2014) 1277-1285.

48. S.Weinberger, What is…Persistent Homology? Notices AMS January 2011 pp. 36-39

49. Daniel Yekutieli. Adjusted Bayesian inference for selected parameters. In: Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74.3 (June 2012) pp. 515-541.

50. D. Yogeshwaran and R.J. Adler, On the topology of random complexes built over stationary point processes. *Annals of Applied Probability* 25, 2015, 3338–3380.

51. D. Yogeshwaran, E. Subag and R.J. Adler, Random geometric complexes in the thermodynamic regime. *Probability Theory and Related Fields*, 2016. In press. (35 pages).

52. Y.Baryshnikov, Harris construction, snakes, barcodes. (in preparation)

53. Y.Baryshnikov, Cyclicity (in preparation)

54. Xiaoying Tian Harris et al. Selective sampling after solving a convex problem. In: arXiv:1609.05609 [math, stat] (Sept. 2016). arXiv: 1609.05609.

Popular articles

1. R.J. Adler, TOPOS, and why you should care about it. Bulletin IMS, 43–2, 2014, 4–5.
2. R.J. Adler, TOPOS: Applied topologists do it with persistence, Bulletin IMS, 43–6, 2014,10–11.
3. R.J. Adler, TOPOS: Pinsky was wrong, Euler was right. Bulletin IMS, 43–8, 2014, 6–7.
4. R.J. Adler, TOPOS: Let’s not make the same mistake twice. Bulletin IMS, 44–

2,2015, 4-5.

Training of Graduate and Postdoctoral Fellows

Omer Bobrowski, a student of Adler, was partly trained on this grant. After a postdoc at Duke, has taken a tenure-track position at the Technion.

Weinberger's student, Katharine Turner, got her Ph.D. and is now a postdoc at EFPL (Lausanne).

Eliran Subag did his masters at the Technion, and moved to Weizmann for Ph.D. studies, and will be receiving his PhD in 2017.

Gregory Naitzat moved to University of Chicago (Statistics) from the Technion. Yogeshwaran Dhandapani, who was a postdoc at Technion moved to the Indian Statistical Institute in Bangalore.

Sunder Ram Krishnan is continuing with his PhD and should finish in 2017.

A recent Stanford PhD, Michael Lesnick, visited the Technion for 3 months, before taking up a postdoctoral position at Princeton (to work with MacPherson). He is now at the Princeton Neuroscience Institute.

Han Wang, a PhD student at UIUC, defended PhD and is now a postdoc at NCSU.

Harish Chintakunta,a postdoc of Baryshnikov's is now at Florida Polytechnic.

Yuriy Mileyko, another of Baryshnikov's postdocs moved to University of Hawaii.

Selected talks and conference organization.

In January 2012, AMS Short Course on *Random Fields and Random Geometry*, was organized by Adler and Taylor at the AMS Annual Meeting, Boston.

Adler coordinated a tutorial on An Introduction to Statistics and Probability for Topologists at the IMA in October 2013, as well as being one of the organizers of a workshop on Topological Data Anaysis which followed the tutorial session. In February 2014, he coorganized the SAMSI workshop on *LDHD: Topological Data Analysis*.

In April, 2015, Adler gave the Annual de Rahm Lecture, at EPFL, Lausanne, Switzerland. (Phase Transitions and Random Topology.)

Adler also was a member of the Scientific Committee of the June 2015 meeting DyToComp (Dynamics, Topology and Computations), in Bedlewo, Poland. In August 2015, at the Stochastic Geometry Workshop, in Poitiers, France, he spoke on Topological Phase Transitions and also gave a course on applied topology. In September 2015, he spoke at the Heilbronn Annual Conference, Bristol, UK.

Adler and Taylor coorganized the October 2013, IMA Tutorial: An Introduction to Statistics and Probability for Topologists, and the October 2013, IMA Workshop on Topological Data Analysis. Co-Organiser. Weinberger spoke at this meeting.

Adler co-organized February 2014: SAMSI workshop on Low Dimensional Structure in High Dimensional Systems: Workshop on Topological Data Analysis. Adler was a member of the scientific committee of Extreme Value Analysis, EVA15, Ann Arbor, Michigan. Adler spoke in November 2014 at the Workshop on Discrete, Computational and Algebraic Topology, Copenhagen. (Pondering Persistence and Extolling Euler.) Finally, in Summer/Fall 2016, Adler was a member of the scientific committee of the *Thematic Semester on Probabilistic Methods in Geometry, Topology, and Spectral Theory*. Centre de Recherches Math´ematiques, Montreal.

Baryshnikov and Weinberger gave plenary talks at ATMCS 6 (British Columbia) in May 2014.

Baryshnikov organized a special semester on applied algebraic topology at ICERM in Fall 2016.

Baryshnikov presented some of the results on random networks at NIST-Bell Labs workshop on Geometry of Networks, at NIST, the MCA special session on Applied Algebraic Topology, and a plenary talk at the SIAM conference on Applied Algebraic Geometry

Baryshnikov, Taylor and Weinberger all spoke at Stochastic Processes and Random Fields: Geometry and Fine properties, at Technion, June 2015

Taylor spoke at Statistical Inference for Large Scale Data, Simon Fraser, April 2015

Taylor spoke at a special Topological Data Analysis workshop at NIPS in December 2012. He was an invited speaker at the European Meeting of Statisticians and participated at the IMA workshop in October 2013.

Taylor and his collaborators gave several talks at the Joint Statistical Meetings in Boston in August 2014.

Taylor also gave the Berkeley-Stanford Colloquium, Berkeley, April 2015 and at JSM 2015 he organized an ISM invited session on Post-Selection Inference" at JSM2015. One of speakers was student Joshua Loftus, speaking on their joint work. He (Talyor) also gave invited talk in the session on “Modern Inferential Methods for Big Data Analysis".

Weinberger gave a plenary talk at the Applied Algebraic Topology meeting in Bedlewo. He gave the “Frontiers of Mathematics” lecture series at Texas A&M; one of the lectures featuring ideas related to property testing and its connections to both pure and applied problems. He lectured three times at IMA during 2013-14, and visited ICERM four times in Fall 2016.

Weinberger organized a conference “Geometric Methods in Data Analysis” in May 2015 at the Stevanovich Center for Financial Mathematics (Chicago). Adler and Baryshnikov were invited speakers. At June, 2015, DyToComp (Dynamics, Topology and Computations), Bedlewo, Poland. Adler was a member of the Scientific Committee. Weinberger was an invited speaker.

Weinberger spoke at the joint IAS-Penn-Rutgers seminar on applied topology and gave a colloquium at Yale (on quantitative topology, which is a theme that overlaps this project). He also gave a lecture in the Simons Science Series at the Simons foundation in New York. Both of these will took place in November 2014. In February, he gave the applied math colloquium at Stanford.

Weinberger organized a conference “Geometric Methods in Data Analysis” in May 2015 at the Stevanovich Center for Financial Mathematics (Chicago). Baryshnikov and Bobrowski both were among the invited speakers.

Finally, we are very happy to report on Adler's four pieces for the IMS, introducing the ideas of applied topology to a very broad audience of statisticians.

Honors and awards.

Adler was invited to give a plenary Special Invited Lecture at the European Meeting of Statisticians in Budapest, July 2013 and was awarded a prestigious European Research Council Advanced grant. He was awarded the 2014 Henry Taub Prize for Academic Excellence at Technion, and gave the 2015 de Rham lecture of the Swiss Doctoral Programme, and the 2015 Heilbronn lecture. Finally, he was invited to give a Plenary lecture at the 2016 British Mathematical Colloquium.

Jonathan Taylor gave an invited lecture at the Bernoulli World Congress 2016, Toronto, Canada. 2016 and the Scandinavian Journal of Statistics invited talk: Selective inference in regression. At NORDSTAT 2016, Copenhagen, Denmark. 2016.

Shmuel Weinberger was inducted in 2012 into the inaugural class of Fellows of the American Mathematical Society. In 2013 he became a Fellow of the American Association for the Advancement of Science. In 2015, Weinberger was appointed the Andrew MacLeish Distinguished Professor of Mathematics at the University of Chicago. He gave the Frontiers in Mathematics lecture series at Texas A&M in 2013, the MINT Distinguished lectures at Tel Aviv University in November 2015, and was invited to give the 2017 Minerva lectures at Princeton University, an invited lecture at the 2017 Mathematical Congress of the Americas and a plenary lecture at the tri-annual meeting of FoCM in Madrid.

The graduate student, Turner received the 2013 Stevanovich Center for Financial Math Fellowship for her work on the Persistent Homology transform and its application to evolutionary biology data,

AFOSR Deliverables Submission Survey

Response ID:7433 Data

1.

Report Type

Final Report

Primary Contact Email
Contact email if there is a problem with the report.

schmuel@math.uchicago.edu

Primary Contact Phone Number
Contact phone number if there is a problem with the report

773-702-7349

Organization / Institution name

The University of Chicago

Grant/Contract Title
The full title of the funded effort.

SATA: STOCHASTIC ALGEBRAIC TOPOLOGY AND APPLICATIONS

Grant/Contract Number
AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-11-1-0216

Principal Investigator Name
The full name of the principal investigator on the grant or contract.

Schmuel Weinberger

Program Officer
The AFOSR Program Officer currently assigned to the award

Dr. Tristan Nguyen

Reporting Period Start Date

09/30/2011

Reporting Period End Date

09/29/2016

Abstract

This project was devoted mainly to applications of topology primarily to data analysis, but also to some engineering (e.g. control) problems. Because of noise and uncertain environments, stochasticity is an important element. Topological invariants are robust to some errors in the bulk, but can frequently be highly sensitive to outliers. The work done in this project concerns the amount of data necessary to solve topological inference, even free of noise, and also the nature of errors caused by noise: Different kinds of tail behavior have very different implications, and heavy tails are shown to have severe implications for these methods. Also studied is how much data is necessary to compute topological invariants robustly as a complexity theoretical problem and also as an analysis of algorithms problem, and under what kinds of conditions of local featurelessness of the data (sometimes called a condition number or feature size)? The study of critical points is applied to using these methods for inference within machine learning, and the topology of configuration spaces is applied to control problems. Finally, several of the papers studied nontraditional integrals (Euler integration) which are related to the Gaussian Kinematic Formula, and have earlier been used for target enumeration, and which now seem to be highly computable and mathematically tractable invariants that define signatures that can be of use in machine learning settings

Distribution Statement

This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

sf0298-Weinberger.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

SATA+Final+report+2016-Weinberger.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

R.J. Adler, E. Subag and J.E. Taylor, Rotation and scale space random fields and the Gaussian kinematic formula. Annals of Statistics, 40, 2012, 2910–2942.

R.J. Adler, G. Samorodnitsky and J.E. Taylor, High level excursion set geometry for non-Gaussian infinitely divisible random fields. Annals of Probability, 41, 2013, 134–169.

R.J. Adler, E. Moldavskaya and G. Samorodnitsky, On the existence of paths between two points in high level excursion sets of Gaussian random fields. Annals of Probability, 42, 2014, 1020–1053.

R.J. Adler, O. Bobrowski and S. Weinberger, Crackle: The homology of noise. Discrete and Computational Geometry, 52, 2014, 680–704.

O. Bobrowski and R.J. Adler, Distance functions, critical points, and topology for some random complexes. Homology, Homotopy and Applications, 16, 2014, 311–344.

R.J. Adler and G. Samorodnitsky, Climbing down Gaussian peaks. Annals of Probability, 2016. In press. (33 pages)

R.J. Adler, S.R. Krishnan, J.E. Taylor and S. Weinberger, Convergence of the reach for a sequence of Gaussian-embedded manifolds. Submitted for publication. (50 pages) arXiv:1503.01733

M. Arnold, Yu. Baryshnikov, Yu. Mileyko, Typical representatives of free homotopy classes in a multi-punctured plane, 2015, submitted.

Yu. Baryshnikov, J. Duda, W. Szpankowski, Types of Markov Fields and Tilings, IEEE Trans. on Information Theory, 2016.

Yu. Baryshnikov, V. Blumen, K. Kim, V. Zharnitsky Billiard dynamics of bouncing dumbbell, Physica D: Nonlinear Phenomena 269, 21-27, 2015.

Yu. Baryshnikov, B. Shapiro, How to Run a Centipede: a Topological Perspective In: Geometric Control Theory and sub-Riemannian Geometry, Springer INdAM Series 5, 37-51, 2014.

Xiaoying, Tian, Harris et al. Selective sampling after solving a convex problem". arXiv:1609.05609 [math, stat] (Sept. 2016). arXiv: 1609.05609.

Baryshnikov, Yuliy; Chen, Cheng; Wang, Han,A design of hybrid feedback stabilization on 1D coverage with topological perspectives,"American Control Conference (ACC), 2016,5154-5160,2016,American Automatic Control Council (AACC)

Y. Baryshnikov, R. Ghrist, M. Wright Hadwiger's Theorem for Definable Functions, Adv. Math. 245, 573-586, 2014.

Y. Baryshnikov, Liberzon, Daniel,Robust stability conditions for switched linear systems: Commutator bounds and the Łojasiewicz inequality,52nd IEEE Conference on Decision and Control,722-726,2013,IEEE

O.Bobrowski, S.Mukherjee, and J.Taylor, Topological Consistency via Kernel Estimation, arXiv:1407.5272 [math, stat] (July 2014) arXiv: 1407.5272. to appear in Bernoulli

O.Bobrowski and S.Weinberger, On the vanishing of homology in Random Cech complexes. Random

Structures and Algorithms, published online 7 November 2016

Yunjin Choi, Jonathan Taylor, and Robert Tibshirani. Selecting the number of principal components: estimation of the true rank of a noisy matrix. arXiv:1410.8260 [stat], October 2014. arXiv: 1410.8260.

A.Dranishnikov, S.Ferry, and S.Weinberger, An infinite dimensional phenomenon in finite dimensional topology (submitted) arXiv:math/0611004

William Fithian, Dennis Sun, and Jonathan Taylor. Optimal Inference After Model Selection. arXiv:1410.2597 [math, stat], October 2014. arXiv: 1410.2597.

Max Grazier G. Sell et al. Sequential selection procedures and false discovery rate control". en. In: Journal of the Royal Statistical Society: Series B (Statistical Methodology 78.2 (Mar. 2016), pp. 423-444.

S.R. Krishnan, J.E. Taylor and R.J. Adler, The intrinsic geometry of some random manifolds. Electronic Communications in Probability, in press. (13 pages)

Jason D Lee, Jonathan E Taylor, Exact Post Model Selection Inference for Marginal Screening (submitted) arXiv:1402.5596

Jason D. Lee, Dennis L. Sun, Yuekai Sun, Jonathan E. , "Exact post-selection inference with the lasso". Annals Statistics 44.3 (June 2016), pp. 907-927.

Stephen Reid, Jonathan Taylor, Robert Tibshirani, Post-selection point and interval estimation of signal sizes in Gaussian samples, In: Journal of the American Statistical Association (Oct. 2016),

Xiaoying Tian, Joshua R. Loftus, and Jonathan E. Taylor. Selective inference with unknown variance via the square-root LASSO. arXiv:1504.08031 [math, stat], April 2015. arXiv: 1504.08031.

Joshua R. Loftus and Jonathan E. Taylor. A significance test for forward stepwise model selection. arXiv:1405.3920 [stat], May 2014. arXiv: 1405.3920.

E Munch, K Turner, P Bendich, S Mukherjee, J Mattingly, J Harer Probabilistic Fréchet means for time varying persistence diagrams Electronic Journal of Statistics 9, 1173-1204

G. Naitzat and R.J. Adler, A central limit theorem for the Euler integral of a Gaussian random field. Stochastic Processes and its Applications, 2016. In press. (32 pages)

T. Owada and R.J. Adler, Limit theorem for point processes under geometric constraints (and topological crackle) Annals of Probability, 2016. In press. (54 pages)

T. Owada, Functional central limit theorem for subgraph counting processes, Annals of Probability, 2016. In press. (35 pages)

Snigdha Panigrahi, Jonathan Taylor, and Asaf Weinstein. Bayesian Post-Selection Inference in the Linear Model". In: arXiv:1605.08824 [stat] (May 2016). arXiv: 1605.08824.

Jonathan Taylor and Robert Tibshirani. Post-selection inference for L1-penalized likelihood models. In: Canadian Journal of Statistics To appear (Feb. 2016).

Xiaoying Tian, Nan Bi, and Jonathan Taylor. MAGIC: a general, powerful and tractable method for selective inference". In: arXiv:1607.02630 [math, stat] (July 2016). arXiv: 1607.02630.

G. Thoppe, D. Yogeshwaran, and R.J. Adler, On the evolution of topology in dynamic clique complexes. Advances in Applied Probability, 48, 2016. In press. (33 pages)

Xiaoying Tian and Jonathan E. Taylor. Selective inference with a randomized response". In: arXiv:1507.06739 [math, stat] (July 2015). arXiv: 1507.06739.

Ryan J. Tibshirani et al. Exact Post-Selection Inference for Sequential Regression Procedures". In: Journal of the American Statistical Association 111.514 (Apr. 2016), pp. 600-620.

K.Turner, Cone fields and topological sampling in manifolds with bounded curvature, Journal of FoCM 13 (2013), no. 6, 913–933.

K Turner, Y Mileyko, S Mukherjee, J Harer , Fréchet means for distributions of persistence diagrams Discrete & Computational Geometry 52 (1), 44-70

Katharine Turner, Sayan Mukherjee, Doug M Boyer, Persistent homology transform for modeling shapes and surfaces, Information and Inference, iau 11 2014

Wang, Han, Chen, Cheng, Baryshnikov, Yuliy, A Topological Perspective on Cycling Robots for Full Tree Coverage, Algorithmic Foundations of Robotics XI, 659-675, 2015,Springer

S.Weinberger, The complexity of some basic topological inference problems, Journal of FoCM, 14 (2014) 1277-1285.

S.Weinberger, What is…Persistent Homology? Notices AMS January 2011 pp. 36-39

Daniel Yekutieli. Adjusted Bayesian inference for selected parameters. In: Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74.3 (June 2012) pp. 515-541.
D. Yogeshwaran and R.J. Adler, On the topology of random complexes built over stationary point processes. Annals of Applied Probability 25, 2015, 3338–3380.
D. Yogeshwaran, E. Subag and R.J. Adler, Random geometric complexes in the thermodynamic regime. Probability Theory and Related Fields, 2016. In press. (35 pages).
Xiaoying Tian Harris et al. Selective sampling after solving a convex problem. In: arXiv:1609.05609 [math, stat] (Sept. 2016). arXiv: 1609.05609.

R.J. Adler, TOPOS, and why you should care about it. Bulletin IMS, 43–2, 2014, 4–5.
R.J. Adler, TOPOS: Applied topologists do it with persistence, Bulletin IMS, 43– 6, 2014,10–11.
R.J. Adler, TOPOS: Pinsky was wrong, Euler was right. Bulletin IMS, 43–8, 2014, 6–7.
R.J. Adler, TOPOS: Let's not make the same mistake twice. Bulletin IMS, 44–2,2015, 4-5.

New discoveries, inventions, or patent disclosures:
Do you have any discoveries, inventions, or patent disclosures to report for this period?

No

Please describe and include any notable dates

Do you plan to pursue a claim for personal or organizational intellectual property?

Changes in research objectives (if any):

As the work developed, we discovered new opportunities including the topological study of multi scale data, and the use of the short bars in a persistence diagram for possible inference

Change in AFOSR Program Officer, if any:

Robert Bonneau
Tristan Nguyen

Extensions granted or milestones slipped, if any:

N/A

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Dec 29, 2016 21:44:54 Success: Email Sent to: schmuel@math.uchicago.edu